

Homework #2: Crawling

Due: 1/23, Wednesday

100 points

(TA handling this homework: Binh Vu binhlvu@usc.edu)

In this homework, you are asked to crawl the descriptions of courses in UCLA. The seed URL is: <https://www.registrar.ucla.edu/Academics/Course-Descriptions>. The following is a screenshot of the seed web site. For each subject area (e.g., Aerospace Studies), the crawler needs to follow its link and download the page that contains the descriptions of courses in that subject area. The page should be stored using the name of the subject area, e.g., "Aerospace Studies.html".

Note: you are free to use any crawler program (e.g., Scrapy that you have seen in class).

Submission:

- **src:** Folder contains your crawler code
- **data.zip:** All the pages you crawl in a zip file.

COURSE DESCRIPTIONS

Official, approved undergraduate and graduate course descriptions are updated daily.

Course descriptions are approved by the Academic Senate and maintained by the Registrar's Office. To browse courses by subject area, click on the subject name. To search courses, enter keyword(s) in the field and click the search button.

For more information on the organization of courses, visit the [course numbering and description guide](#).

To find past course descriptions, see the [UCLA General Catalog archive](#).

<input type="text" value="Search course descriptions and titles"/>	<input type="button" value="Search Classes"/>
--	---

Browse By Subject Area

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	R	S	T	U	V
W	Y																			

A	Aerospace Studies	Archaeology
	African American Studies	Architecture and Urban Design
	African Studies	Armenian
	Afrikaans	Art
	American Indian Studies	Art History
	American Sign Language	Arts and Architecture
	Ancient Near East	Arts Education
	Anesthesiology	Asian
	Anthropology	Asian American Studies